DECLARATION OF ERIC BRENNER IN OPPOSITION TO PLAINTIFFS' MOTION IN LIMINE TO EXCLUDE THE LITIGATION SURVEY BY DEFENDANTS' EXPERT KEVIN LANE KELLER, PH.D.

EXHIBIT 8

Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

The definition of the relevant population is crucial because there may be systematic differences in the responses of members of the population and nonmembers. For example, consumers who are prospective purchasers may know more about the product category than consumers who are not considering making a purchase.

The universe must be defined carefully. For example, a commercial for a toy or breakfast cereal may be aimed at children, who in turn influence their parents' purchases. If a survey assessing the commercial's tendency to mislead were conducted based on a sample from the target population of prospective and actual adult purchasers, it would exclude a crucial relevant population. The appropriate population in this instance would include children as well as parents.[78]

B. Did the Sampling Frame Approximate the Population?

The target population consists of all the individuals or units that the researcher would like to study. The sampling frame is the source (or sources) from which the sample actually is drawn. The surveyor's job generally is easier if a complete list of every eligible member of the population is available (e.g., all plaintiffs in a discovery survey), so that the sampling frame lists the identity of all members of the target population. Frequently, however, the target population includes members who are inaccessible or who cannot be identified in advance. As a result, reasonable compromises are sometimes required in developing the sampling frame. The survey report should contain (1) a description of the target population, (2) a description of the sampling frame from which the sample is to be drawn, (3) a discussion of the difference between the target population and the sampling frame, and, importantly, (4) an evaluation of the likely consequences of that difference.

A survey that provides information about a wholly irrelevant population is itself irrelevant.[79] Courts are likely to exclude the survey or accord it little

trial, that the defendants were guilty); *see also* People v. Venegas, 31 Cal. Rptr. 2d 114, 117 (Cal. Ct. App. 1994) (change of venue denied because defendant failed to show that the defendant would face a less hostile jury in a different court).

78. *See, e.g.*, Warner Bros., Inc. v. Gay Toys, Inc., 658 F.2d 76 (2d Cir. 1981) (surveying children users of the product rather than parent purchasers). Children and some other populations create special challenges for researchers. For example, very young children should not be asked about sponsorship or licensing, concepts that are foreign to them. Concepts, as well as wording, should be age appropriate.

79. A survey aimed at assessing how persons in the trade respond to an advertisement should be conducted on a sample of persons in the trade and not on a sample of consumers. *See* Home Box Office v. Showtime/The Movie Channel, 665 F. Supp. 1079, 1083 (S.D.N.Y.), *aff'd in part and vacated in part*, 832 F.2d 1311 (2d Cir. 1987); J & J Snack Food Corp. v. Earthgrains Co., 220 F. Supp. 2d 358, 371–72 (N.J. 2002). *But see* Lon Tai Shing Co. v. Koch + Lowy, No. 90-C4464, 1990 U.S. Dist. LEXIS 19123, at *50 (S.D.N.Y. Dec. 14, 1990), in which the judge was willing to find likelihood of consumer confusion from a survey of lighting store salespersons questioned by a survey researcher posing as a customer. The court was persuaded that the salespersons who were misstating the source

weight.[80] Thus, when the plaintiff submitted the results of a survey to prove that the green color of its fishing rod had acquired a secondary meaning, the court gave the survey little weight in part because the survey solicited the views of fishing rod dealers rather than consumers.[81] More commonly, however, the sampling frame and the target population have some overlap, but the overlap is imperfect: The sampling frame excludes part of the target population, that is, it is underinclusive, or the sampling frame includes individuals who are not members of the target population, that is, it is overinclusive relative to the target population. Coverage error is the term used to describe inconsistencies between a sampling frame and a target population. If the coverage is underinclusive, the survey's value depends on the proportion of the target population that has been excluded from the sampling frame and the extent to which the excluded population is likely to respond differently from the included population. Thus, a survey of spectators and participants at running events would be sampling a sophisticated subset of those likely to purchase running shoes. Because this subset probably would consist of the consumers most knowledgeable about the trade dress used by companies that sell running shoes, a survey based on this sampling frame would be likely to substantially overrepresent the strength of a particular design as a trademark, and the extent of that overrepresentation would be unknown and not susceptible to any reasonable estimation.[82]

Similarly, in a survey designed to project demand for cellular phones, the assumption that businesses would be the primary users of cellular service led surveyors to exclude potential nonbusiness users from the survey. The Federal Communications Commission (FCC) found the assumption unwarranted and concluded that the research was flawed, in part because of this underinclusive coverage.[83] With the growth in individual cell phone use over time, noncoverage error would be an even greater problem for this survey today.

of the lamp, whether consciously or not, must have believed reasonably that the consuming public would be likely to rely on the salespersons' inaccurate statements about the name of the company that manufactured the lamp they were selling.

80. *See* Wells Fargo & Co. v. WhenU.com, Inc., 293 F. Supp. 2d 734 (E.D. Mich. 2003).

81. *See* R.L. Winston Rod Co. v. Sage Mfg. Co., 838 F. Supp. 1396, 1401–02 (D. Mont. 1993).

82. *See* Brooks Shoe Mfg. Co. v. Suave Shoe Corp., 533 F. Supp. 75, 80 (S.D. Fla. 1981), *aff'd*, 716 F.2d 854 (11th Cir. 1983); *see also* Hodgdon Power Co. v. Alliant Techsystems, Inc., 512 F. Supp. 2d 1178 (D. Kan. 2007) (excluding survey on gunpowder brands distributed at plaintiff's promotional booth at a shooting tournament); Winning Ways, Inc. v. Holloway Sportswear, Inc., 913 F. Supp. 1454, 1467 (D. Kan. 1996) (survey flawed in failing to include sporting goods customers who constituted a major portion of customers). *But see* Thomas & Betts Corp. v. Panduit Corp., 138 F.3d 277, 294–95 (7th Cir. 1998) (survey of store personnel admissible because relevant market included both distributors and ultimate purchasers).

83. *See* Gencom, Inc., 56 Rad. Reg. 2d (P&F) 1597, 1604 (1984). This position was affirmed on appeal. *See* Gencom, Inc. v. FCC, 832 F.2d 171, 186 (D.C. Cir. 1987); *see also* Beacon Mut. Ins. Co. v. Onebeacon Ins. Corp, 376 F. Supp. 2d 251, 261 (D.R.I. 2005) (sample included only defendant's insurance agents and lack of confusion among those agents was "nonstartling").

In some cases, it is difficult to determine whether a sampling frame that omits some members of the population distorts the results of the survey and, if so, the extent and likely direction of the bias. For example, a trademark survey was designed to test the likelihood of confusing an analgesic currently on the market with a new product that was similar in appearance.[84] The plaintiff's survey included only respondents who had used the plaintiff's analgesic, and the court found that the target population should have included users of other analgesics, "so that the full range of potential customers for whom plaintiff and defendants would compete could be studied."[85] In this instance, it is unclear whether users of the plaintiff's product would be more or less likely to be confused than users of the defendants' product or users of a third analgesic.[86]

An overinclusive sampling frame generally presents less of a problem for interpretation than does an underinclusive sampling frame.[87] If the survey expert can demonstrate that a sufficiently large (and representative) subset of respondents in the survey was drawn from the appropriate sampling frame, the responses obtained from that subset can be examined, and inferences about the relevant population can be drawn based on that subset.[88] If the relevant subset cannot be identified, however, an overbroad sampling frame will reduce the value of the survey.[89] If the sampling frame does not include important groups in the target population, there is generally no way to know how the unrepresented members of the target population would have responded.[90]

84. *See* American Home Prods. Corp. v. Barr Lab., Inc., 656 F. Supp. 1058 (D.N.J.), *aff'd*, 834 F.2d 368 (3d Cir. 1987).

85. *Id*. at 1070.

86. *See also* Craig v. Boren, 429 U.S. 190 (1976).

87. *See* Schwab v. Philip Morris USA, Inc. 449 F. Supp. 2d 992, 1134–35 (E.D.N.Y. 2006) ("Studies evaluating broadly the beliefs of low tar smokers generally are relevant to the beliefs of "light" smokers more specifically.").

88. *See* National Football League Props. Inc. v. Wichita Falls Sportswear, Inc. 532 F. Supp. 651, 657–58 (W.D. Wash. 1982).

89. *See* Leelanau Wine Cellars, Ltd. v. Black & Red, Inc., 502 F.3d 504, 518 (6th Cir. 2007) (lower court was correct in giving little weight to survey with overbroad universe); Big Dog Motorcycles, L.L.C. v. Big Dog Holdings, Inc., 402 F. Supp. 2d 1312, 1334 (D. Kan. 2005) (universe composed of prospective purchasers of all t-shirts and caps overinclusive for evaluating reactions of buyers likely to purchase merchandise at motorcycle dealerships). *See also* Schieffelin & Co. v. Jack Co. of Boca, 850 F. Supp. 232, 246 (S.D.N.Y. 1994).

90. *See, e.g.*, Amstar Corp. v. Domino's Pizza, Inc., 615 F.2d 252, 263–64 (5th Cir. 1980) (court found both plaintiff's and defendant's surveys substantially defective for a systematic failure to include parts of the relevant population); Scott Fetzer Co. v. House of Vacuums, Inc., 381 F.3d 477 (5th Cir. 2004) (universe drawn from plaintiff's customer list underinclusive and likely to differ in their familiarity with plaintiff's marketing and distribution techniques).

same type of respondents who would be eligible to participate in the full-scale survey. The interviewers observe the respondents for any difficulties they may have with the questions and probe for the source of any such difficulties so that the questions can be rephrased if confusion or other difficulties arise.[130] Attorneys who commission surveys for litigation sometimes are reluctant to approve pilot work or to reveal that pilot work has taken place because they are concerned that if a pretest leads to revised wording of the questions, the trier of fact may believe that the survey has been manipulated and is biased or unfair. A more appropriate reaction is to recognize that pilot work is a standard and valuable way to improve the quality of a survey[131] and to anticipate that it often results in word changes that increase clarity and correct misunderstandings. Thus, changes may indicate informed survey construction rather than flawed survey design.[132]

B. Were Some Respondents Likely to Have No Opinion? If So, What Steps Were Taken to Reduce Guessing?

Some survey respondents may have no opinion on an issue under investigation, either because they have never thought about it before or because the question mistakenly assumes a familiarity with the issue. For example, survey respondents may not have noticed that the commercial they are being questioned about guaranteed the quality of the product being advertised and thus they may have no opinion on the kind of guarantee it indicated. Likewise, in an employee survey, respondents may not be familiar with the parental leave policy at their company and thus may have no opinion on whether they would consider taking advantage of the parental leave policy if they became parents. The following three alternative question structures will affect how those respondents answer and how their responses are counted.

First, the survey can ask all respondents to answer the question (e.g., "Did you understand the guarantee offered by Clover to be a 1-year guarantee, a 60-day guarantee, or a 30-day guarantee?"). Faced with a direct question, particularly one that provides response alternatives, the respondent obligingly may supply an

130. Methods for testing respondent understanding include concurrent and retrospective think-alouds, in which respondents describe their thinking as they arrive at, or after they have arrived at, an answer, and paraphrasing (asking respondents to restate the question in their own words). Tourangeau et al., *supra* note 113, at 326–27; *see also* Methods for Testing and Evaluating Survey Questionnaires (Stanley Presser et al. eds., 2004).

131. *See* OMB Standards and Guidelines for Statistical Survey, *supra* note 110, Standard 1.4, Pretesting Survey Systems (specifying that to ensure that all components of a survey function as intended, pretests of survey components should be conducted unless those components have previously been successfully fielded); American Association for Public Opinion Research, Best Practices (2011) ("Because it is rarely possible to foresee all the potential misunderstandings or biasing effects of different questions or procedures, it is vital for a well-designed survey operation to include provision for a pretest.").

132. *See infra* Section VII.B for a discussion of obligations to disclose pilot work.

answer even if (in this example) the respondent did not notice the guarantee (or is unfamiliar with the parental leave policy). Such answers will reflect only what the respondent can glean from the question, or they may reflect pure guessing. The imprecision introduced by this approach will increase with the proportion of respondents who are unfamiliar with the topic at issue.

Second, the survey can use a quasi-filter question to reduce guessing by providing "don't know" or "no opinion" options as part of the question (e.g., "Did you understand the guarantee offered by Clover to be for more than a year, a year, or less than a year, or don't you have an opinion?").[133] By signaling to the respondent that it is appropriate not to have an opinion, the question reduces the demand for an answer and, as a result, the inclination to hazard a guess just to comply. Respondents are more likely to choose a "no opinion" option if it is mentioned explicitly by the interviewer than if it is merely accepted when the respondent spontaneously offers it as a response. The consequence of this change in format is substantial. Studies indicate that, although the relative distribution of the respondents selecting the *listed* choices is unlikely to change dramatically, presentation of an explicit "don't know" or "no opinion" alternative commonly leads to a 20% to 25% increase in the proportion of respondents selecting that response.[134]

Finally, the survey can include full-filter questions, that is, questions that lay the groundwork for the substantive question by first asking the respondent if he or she has an opinion about the issue or happened to notice the feature that the interviewer is preparing to ask about (e.g., "Based on the commercial you just saw, do you have an opinion about how long Clover stated or implied that its guarantee lasts?").[135] The interviewer then asks the substantive question only of those respondents who have indicated that they have an opinion on the issue.

Which of these three approaches is used and the way it is used can affect the rate of "no opinion" responses that the substantive question will evoke.[136] Respondents are more likely to say that they do not have an opinion on an issue if a full filter is used than if a quasi-filter is used.[137] However, in maximizing respondent expressions of "no opinion," full filters may produce an underreporting of opinions. There is some evidence that full-filter questions discourage respondents who actually have opinions from offering them by conveying the implicit suggestion that respondents can avoid difficult followup questions by saying that they have no opinion.[138]

133. Norbert Schwarz & Hans-Jürgen Hippler, *Response Alternatives: The Impact of Their Choice and Presentation Order*, *in* Measurement Errors in Surveys 41, 45–46 (Paul P. Biemer et al. eds., 1991).

134. Howard Schuman & Stanley Presser, Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context 113–46 (1981).

135. *See, e.g.*, Johnson & Johnson–Merck Consumer Pharmas. Co. v. SmithKline Beecham Corp., 960 F.2d 294, 299 (2d Cir. 1992).

136. Considerable research has been conducted on the effects of filters. For a review, see George F. Bishop et al., *Effects of Filter Questions in Public Opinion Surveys*, 47 Pub. Op. Q. 528 (1983).

137. Schwarz & Hippler, *supra* note 133, at 45–46.

138. *Id.* at 46.

In general, then, a survey that uses full filters provides a conservative estimate of the number of respondents holding an opinion, while a survey that uses neither full filters nor quasi-filters may overestimate the number of respondents with opinions, if some respondents offering opinions are guessing. The strategy of including a "no opinion" or "don't know" response as a quasi-filter avoids both of these extremes. Thus, rather than asking, "Based on the commercial, do you believe that the two products are made in the same way, or are they made differently?"[139] or prefacing the question with a preliminary, "Do you have an opinion, based on the commercial, concerning the way that the two products are made?" the question could be phrased, "Based on the commercial, do you believe that the two products are made in the same way, or that they are made differently, or don't you have an opinion about the way they are made?"

Recent research on the effects of including a "don't know" option shows that quasi-filters as well as full filters may discourage a respondent who would be able to provide a meaningful answer from expressing it.[140] The "don't know" option provides a cue that it is acceptable to avoid the work of trying to provide a more substantive response. Respondents are particularly likely to be attracted to a "don't know" option when the question is difficult to understand or the respondent is not strongly motivated to carefully report an opinion.[141] One solution that some survey researchers use is to provide respondents with a general instruction not to guess at the beginning of an interview, rather than supplying a "don't know" or "no opinion" option as part of the options attached to each question.[142] Another approach is to eliminate the "don't know" option and to add followup questions that measure the strength of the respondent's opinion.[143]

C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified?

The questions that make up a survey instrument may be open-ended, closed-ended, or a combination of both. Open-ended questions require the respondent to formulate and express an answer in his or her own words (e.g., "What was the main point of the commercial?" "Where did you catch the fish you caught

139. The question in the example without the "no opinion" alternative was based on a question rejected by the court in Coors Brewing Co. v. Anheuser-Busch Cos., 802 F. Supp. 965, 972–73 (S.D.N.Y. 1992). *See also* Procter & Gamble Pharms., Inc. v. Hoffmann-La Roche, Inc., 2006 U.S. Dist. LEXIS 64363 (S.D.N.Y. Sept. 6, 2006).

140. Jon A. Krosnick et al., *The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or Invitation to Satisfice*? 66 Pub. Op. Q. 371 (2002).

141. Krosnick & Presser, *supra* note 126, at 284.

142. Anheuser-Busch, Inc. v. VIP Prods, LLC, No. 4:08cv0358, 2008 U.S. Dist. LEXIS 82258, at *6 (E.D. Mo. Oct. 16, 2008).

143. Krosnick & Presser, *supra* note 126, at 285.

F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question?

Many surveys are designed not simply to describe attitudes or beliefs or reported behaviors, but to determine the source of those attitudes or beliefs or behaviors. That is, the purpose of the survey is to test a causal proposition. For example, how does a trademark or the content of a commercial affect respondents' perceptions or understanding of a product or commercial? Thus, the question is not merely whether consumers hold inaccurate beliefs about Product A, but whether exposure to the commercial misleads the consumer into thinking that Product A is a superior pain reliever. Yet if consumers already believe, before viewing the commercial, that Product A is a superior pain reliever, a survey that simply records consumers' impressions after they view the commercial may reflect those preexisting beliefs rather than impressions produced by the commercial.

Surveys that merely record consumer impressions have a limited ability to answer questions about the origins of those impressions. The difficulty is that the consumer's response to any question on the survey may be the result of information or misinformation from sources other than the trademark the respondent is being shown or the commercial he or she has just watched.[169] In a trademark survey attempting to show secondary meaning, for example, respondents were shown a picture of the stripes used on Mennen stick deodorant and asked, "[W]hich [brand] would you say uses these stripes on their package?"[170] The court recognized that the high percentage of respondents selecting "Mennen" from an array of brand names may have represented "merely a playback of brand share";[171] that is, respondents asked to give a brand name may guess the one that is most familiar, generally the brand with the largest market share.[172]

Some surveys attempt to reduce the impact of preexisting impressions on respondents' answers by instructing respondents to focus solely on the stimulus as a basis for their answers. Thus, the survey includes a preface (e.g., "based on the commercial you just saw") or directs the respondent's attention to the mark at issue (e.g., "these stripes on the package"). Such efforts are likely to be only partially successful. It is often difficult for respondents to identify accurately the

169. *See, e.g.,* Procter & Gamble Co. v. Ultreo, Inc., 574 F. Supp. 2d. 339, 351–52 (S.D.N.Y. 2008) (survey was unreliable because it failed to control for the effect of preexisting beliefs).

170. Mennen Co. v. Gillette Co., 565 F. Supp. 648, 652 (S.D.N.Y. 1983), *aff'd*, 742 F.2d 1437 (2d Cir. 1984). To demonstrate secondary meaning, "the [c]ourt must determine whether the mark has been so associated in the mind of consumers with the entity that it identifies that the goods sold by that entity are distinguished by the mark or symbol from goods sold by others." *Id.*

171. *Id.*

172. *See also* Upjohn Co. v. American Home Prods. Corp., No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049, at *42–44 (W.D. Mich. Apr. 5, 1996).

source of their impressions.[173] The more routine the idea being examined in the survey (e.g., that the advertised pain reliever is more effective than others on the market; that the mark belongs to the brand with the largest market share), the more likely it is that the respondent's answer is influenced by (1) preexisting impressions; (2) general expectations about what commercials typically say (e.g., the product being advertised is better than its competitors); or (3) guessing, rather than by the actual content of the commercial message or trademark being evaluated.

It is possible to adjust many survey designs so that causal inferences about the effect of a trademark or an allegedly deceptive commercial become clear and unambiguous. By adding one or more appropriate control groups, the survey expert can test directly the influence of the stimulus.[174] In the simplest version of such a survey experiment, respondents are assigned randomly to one of two conditions.[175] For example, respondents assigned to the experimental condition view an allegedly deceptive commercial, and respondents assigned to the control condition either view a commercial that does not contain the allegedly deceptive material or do not view any commercial.[176] Respondents in both the experimental and control groups answer the same set of questions about the allegedly deceptive message. The effect of the commercial's allegedly deceptive message is evaluated by comparing the responses made by the experimental group members with those of the control group members. If 40% of the respondents in the experimental group responded indicating that they received the deceptive message (e.g., the advertised product has fewer calories than its competitor), whereas only 8% of the respondents in the control group gave that response, the difference between 40% and 8% (within the limits of sampling error[177]) can be attributed only to the allegedly deceptive message. Without the control group, it is not possible to determine how much of the 40% is attributable to respondents' preexisting beliefs

173. *See* Richard E. Nisbett & Timothy D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 Psychol. Rev. 231 (1977).

174. *See* Shari S. Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 Law & Hum. Behav. 239, 244–46 (1989); Jacob Jacoby & Constance Small, *Applied Marketing: The FDA Approach to Defining Misleading Advertising*, 39 J. Marketing 65, 68 (1975). *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.A, in this manual.

175. Random assignment should not be confused with random selection. When respondents are assigned randomly to different treatment groups (e.g., respondents in each group watch a different commercial), the procedure ensures that within the limits of sampling error the two groups of respondents will be equivalent except for the different treatments they receive. Respondents selected for a mall intercept study, and not from a probability sample, may be assigned randomly to different treatment groups. Random selection, in contrast, describes the method of selecting a sample of respondents in a probability sample. *See supra* Section III.C.

176. This alternative commercial could be a "tombstone" advertisement that includes only the name of the product or a more elaborate commercial that does not include the claim at issue.

177. For a discussion of sampling error, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.A, in this manual.

or other background noise (e.g., respondents who misunderstand the question or misstate their responses). Both preexisting beliefs and other background noise should have produced similar response levels in the experimental and control groups. In addition, if respondents who viewed the allegedly deceptive commercial respond differently than respondents who viewed the control commercial, the difference cannot be merely the result of a leading question, because both groups answered the same question. The ability to evaluate the effect of the wording of a particular question makes the control group design particularly useful in assessing responses to closed-ended questions,[178] which may encourage guessing or particular responses. Thus, the focus on the response level in a control group design is not on the absolute response level, but on the difference between the response level of the experimental group and that of the control group.[179]

In designing a survey-experiment, the expert should select a stimulus for the control group that shares as many characteristics with the experimental stimulus as possible, with the key exception of the characteristic whose influence is being assessed.[180] Although a survey with an imperfect control group may provide better information than a survey with no control group at all, the choice of an appropriate control group requires some care and should influence the weight that the survey receives. For example, a control stimulus should not be less attractive than the experimental stimulus if the survey is designed to measure how familiar the experimental stimulus is to respondents, because attractiveness may affect perceived familiarity.[181] Nor should the control stimulus share with the experimental stimulus the feature whose impact is being assessed. If, for example, the control stimulus in a case of alleged trademark infringement is itself a likely source of consumer confusion, reactions to the experimental and control stimuli may not

178. The Federal Trade Commission has long recognized the need for some kind of control for closed-ended questions, although it has not specified the type of control that is necessary. *See* Stouffer Foods Corp., 118 F.T.C. 746, No. 9250, 1994 FTC LEXIS 196, at *31 (Sept. 26, 1994).

179. *See, e.g.*, Cytosport, Inc. v. Vital Pharms., Inc., 617 F. Supp. 2d 1051, 1075–76 (E.D. Cal. 2009) (net confusion level of 25.4% obtained by subtracting 26.5% in the control group from 51.9% in the test group).

180. *See, e.g.,* Skechers USA, Inc. v. Vans, Inc., No. CV-07-01703, 2007 WL 4181677, at *8–9 (C.D. Cal. Nov. 20, 2007) (in trade dress infringement case, control stimulus should have retained design elements not at issue); Procter & Gamble Pharms., Inc. v. Hoffman-LaRoche, Inc., No. 06-Civ-0034, 2006 U.S. Dist. LEXIS 64363, at *87 (S.D.N.Y. Sept. 6, 2006) (in false advertising action, disclaimer was inadequate substitute for appropriate control group).

181. *See, e.g.*, Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club L.P., 34 F.3d 410, 415–16 (7th Cir. 1994) (court recognized that the name "Baltimore Horses" was less attractive for a sports team than the name "Baltimore Colts."); *see also* Reed-Union Corp. v. Turtle Wax, Inc., 77 F.3d 909, 912 (7th Cir. 1996) (court noted that one expert's choice of a control brand with a well-known corporate source was less appropriate than the opposing expert's choice of a control brand whose name did not indicate a specific corporate source); Louis Vuitton Malletier v. Dooney & Bourke, Inc., 525 F. Supp. 2d 576, 595 (S.D.N.Y. 2007) (underreporting of background "noise" likely occurred because handbag used as control was quite dissimilar in shape and pattern to both plaintiff and defendant's bags).

differ because both cause respondents to express the same level of confusion.[182] In an extreme case, an inappropriate control may do nothing more than control for the effect of the nature or wording of the survey questions (e.g., acquiescence).[183] That may not be enough to rule out other explanations for different or similar responses to the experimental and control stimuli. Finally, it may sometimes be appropriate to have more than one control group to assess precisely what is causing the response to the experimental stimulus (e.g., in the case of an allegedly deceptive ad, whether it is a misleading graph or a misleading claim by the announcer; or in the case of allegedly infringing trade dress, whether it is the style of the font used or the coloring of the packaging).

Explicit attention to the value of control groups in trademark and deceptive-advertising litigation is a relatively recent phenomenon, but courts have increasingly come to recognize the central role the control group can play in evaluating claims.[184] A LEXIS search using *Lanham Act* and *control group* revealed only 4 federal district court cases before 1991 in which surveys with control groups were discussed, 16 in the 9 years from 1991 to 1999, and 46 in the 9 years between 2000 and 2008, a rate of growth that far exceeds the growth in Lanham Act litigation. In addition, courts in other cases have described or considered surveys using control group designs without labeling the comparison group a control group.[185] Indeed, one reason why cases involving surveys with control groups may be underrepresented in reported cases is that a survey with a control group produces

182. *See, e.g.*, Western Publ'g Co. v. Publications Int'l, Ltd., No. 94-C-6803, 1995 U.S. Dist. LEXIS 5917, at *45 (N.D. Ill. May 2, 1995) (court noted that the control product was "arguably more infringing than" the defendant's product) (emphasis omitted). *See also* Classic Foods Int'l Corp. v. Kettle Foods, Inc., 2006 U.S. Dist. LEXIS 97200 (C.D. Cal. Mar. 2, 2006); McNeil-PPC, Inc. v. Merisant Co., 2004 U.S. Dist. LEXIS 27733 (D.P.R. July 29, 2004).

183. *See* text accompanying note 156, *supra*.

184. *See, e.g.*, SmithKline Beecham Consumer Healthcare, L.P. v. Johnson & Johnson-Merck, 2001 U.S. Dist. LEXIS 7061, at *37 (S.D.N.Y. June 1, 2001) (survey to assess implied falsity of a commercial not probative in the absence of a control group); Consumer American Home Prods. Corp. v. Procter & Gamble Co., 871 F. Supp. 739, 749 (D.N.J. 1994) (discounting survey results based on failure to control for participants' preconceived notions); ConAgra, Inc. v. Geo. A. Hormel & Co., 784 F. Supp. 700, 728 (D. Neb. 1992) ("Since no control was used, the . . . study, standing alone, must be significantly discounted."), *aff'd*, 990 F.2d 368 (8th Cir. 1993).

185. Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club L.P., No. 94727-C, 1994 U.S. Dist. LEXIS 19277, at *10–11 (S.D. Ind. June 27, 1994), *aff'd*, 34 F.3d 410 (7th Cir. 1994). In *Indianapolis Colts*, the district court described a survey conducted by the plaintiff's expert in which half of the interviewees were shown a shirt with the name "Baltimore CFL Colts" on it and half were shown a shirt on which the word "Horses" had been substituted for the word "Colts." *Id.* The court noted that the comparison of reactions to the horse and colt versions of the shirt made it possible "to determine the impact from the use of the word 'Colts.'" *Id.* at *11. *See also* Quality Inns Int'l, Inc. v. McDonald's Corp., 695 F. Supp. 198, 218 (D. Md. 1988) (survey revealed confusion between McDonald's and McSleep, but control survey revealed no confusion between McDonald's and McTavish). *See also* Simon Prop. Group L.P. v. MySimon, Inc., 104 F. Supp. 2d 1033 (S.D. Ind. 2000) (court criticized the survey design based on the absence of a control that could show that results were produced by legally relevant confusion).

less ambiguous findings, which may lead to a resolution before a preliminary injunction hearing or trial occurs.

A less common use of control methodology is a control question. Rather than administering a control stimulus to a separate group of respondents, the survey asks all respondents one or more control questions along with the question about the product or service at issue. In a trademark dispute, for example, a survey indicated that 7.2% of respondents believed that "The Mart" and "K-Mart" were owned by the same individuals. The court found no likelihood of confusion based on survey evidence that 5.7% of the respondents also thought that "The Mart" and "King's Department Store" were owned by the same source.[186]

Similarly, a standard technique used to evaluate whether a brand name is generic is to present survey respondents with a series of product or service names and ask them to indicate in each instance whether they believe the name is a brand name or a common name. By showing that 68% of respondents considered Teflon a brand name (a proportion similar to the 75% of respondents who recognized the acknowledged trademark Jell-O as a brand name, and markedly different from the 13% who thought aspirin was a brand name), the makers of Teflon retained their trademark.[187]

Every measure of opinion or belief in a survey reflects some degree of error. Control groups and, as a second choice, control questions are the most reliable means for assessing response levels against the baseline level of error associated with a particular question.

G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey?

Three primary methods have traditionally been used to collect survey data: (1) in-person interviews, (2) telephone interviews, and (3) mail questionnaires.[188] Recently, in the wake of increasing use of the Internet, researchers have added Web-based surveys to their arsenal of tools. Surveys using in-person and telephone interviews, too, now regularly rely on computerized data collection.[189]

186. S.S. Kresge Co. v. United Factory Outlet, Inc., 598 F.2d 694, 697 (1st Cir. 1979). Note that the aggregate percentages reported here do not reveal how many of the same respondents were confused by both names, an issue that may be relevant in some situations. *See* Joseph L. Gastwirth, *Reference Guide on Survey Research*, 36 Jurimetrics J. 181, 187–88 (1996) (review essay).

187. E.I. du Pont de Nemours & Co. v. Yoshida Int'l, Inc., 393 F. Supp. 502, 526–27 & n.54 (E.D.N.Y. 1975); *see also* Donchez v. Coors Brewing Co., 392 F.3d 1211, 1218 (10th Cir. 2004) (respondents evaluated eight brand and generic names in addition to the disputed name). A similar approach is used in assessing secondary meaning.

188. Methods also may be combined, as when the telephone is used to "screen" for eligible respondents, who then are invited to participate in an in-person interview.

189. Wright & Marsden, *supra* note 1, at 13–14.